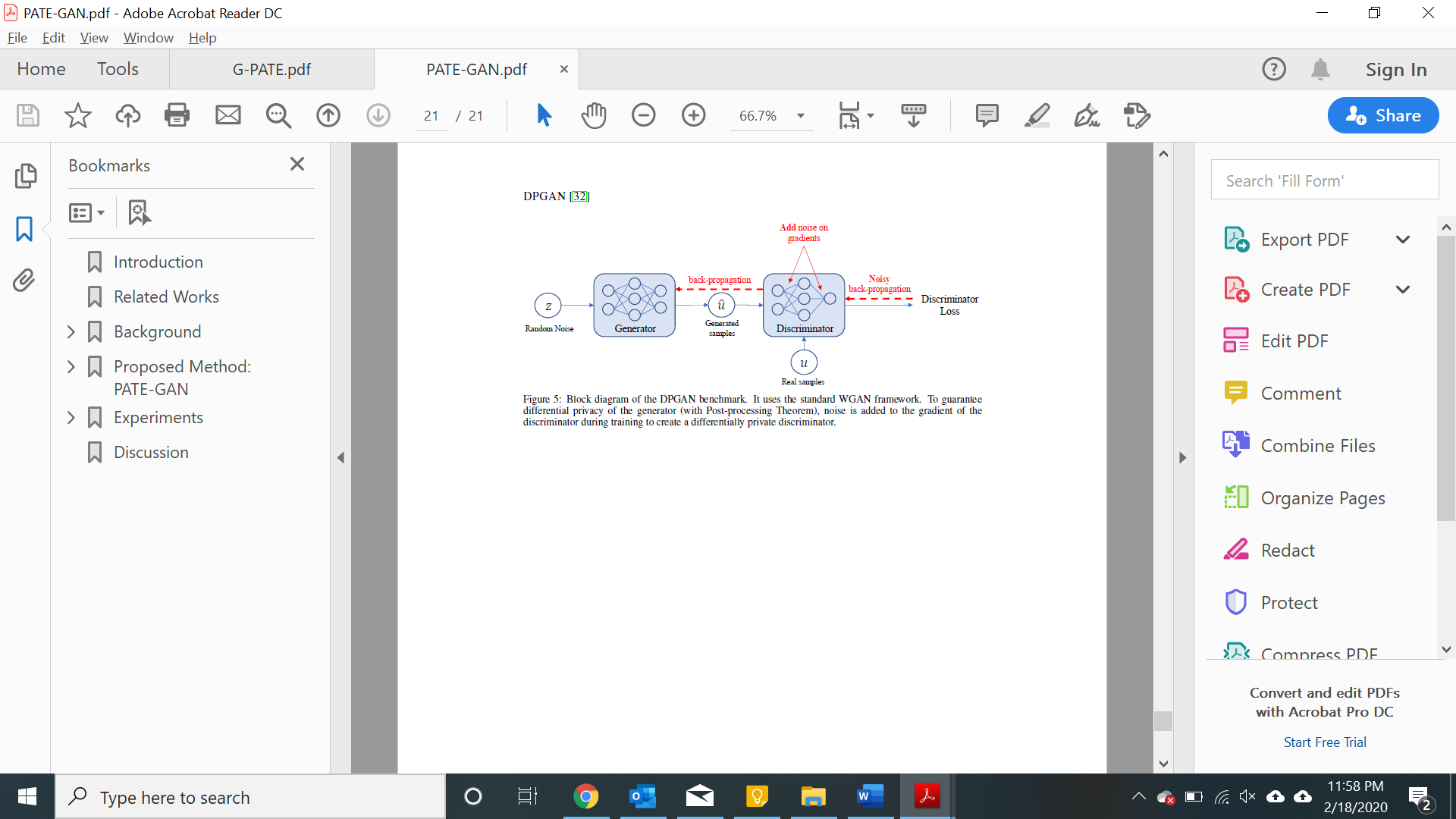
**87% of people can be identified by only 3 fields i.e. Zipcode, gender, birthdate**

**Differential privacy**

1. Add noise to data using Laplace formula
2. Only usable for large dataset since we are injecting noise. Using on small data set will result in inaccurate data

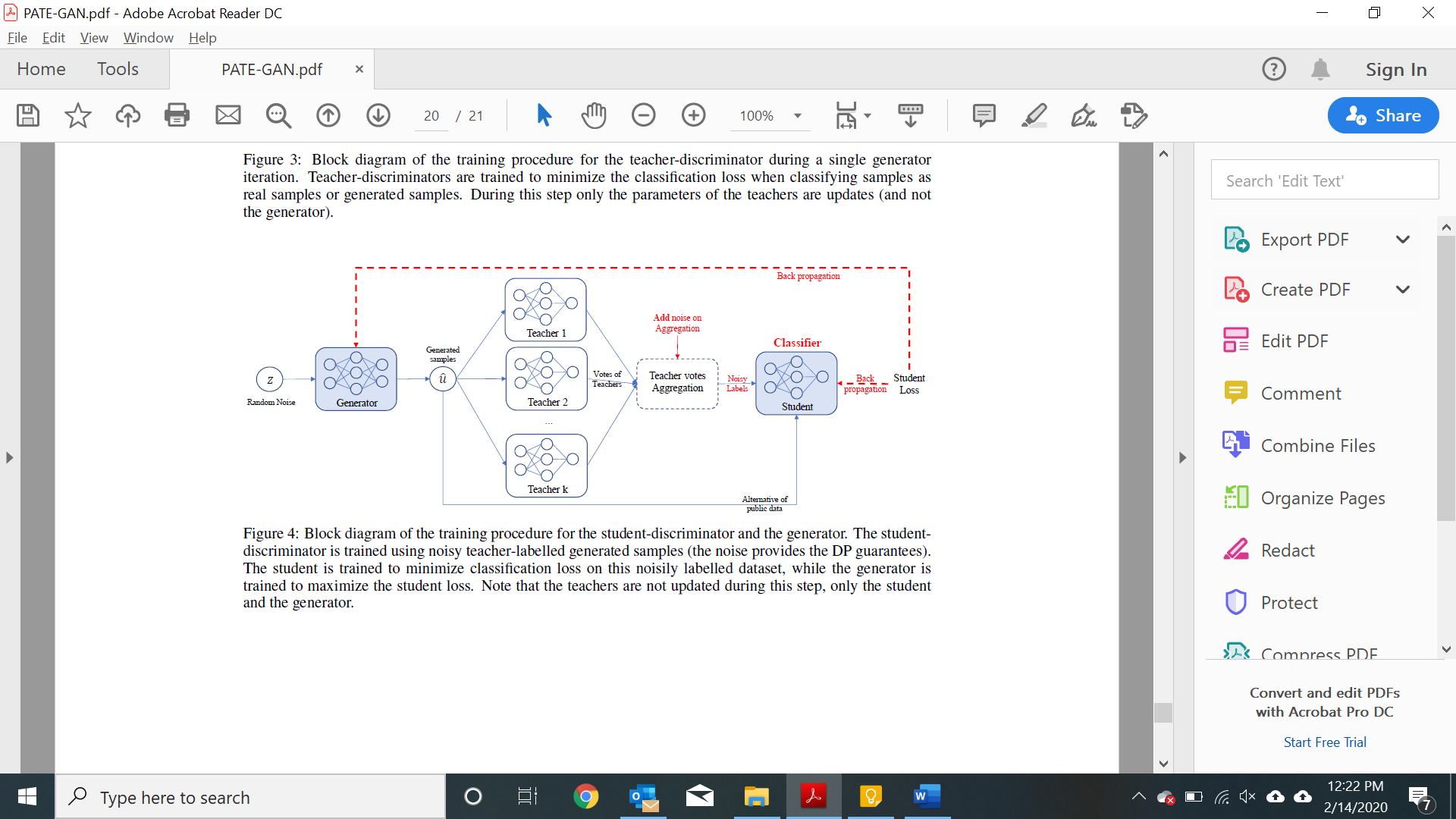
**DP-GAN**

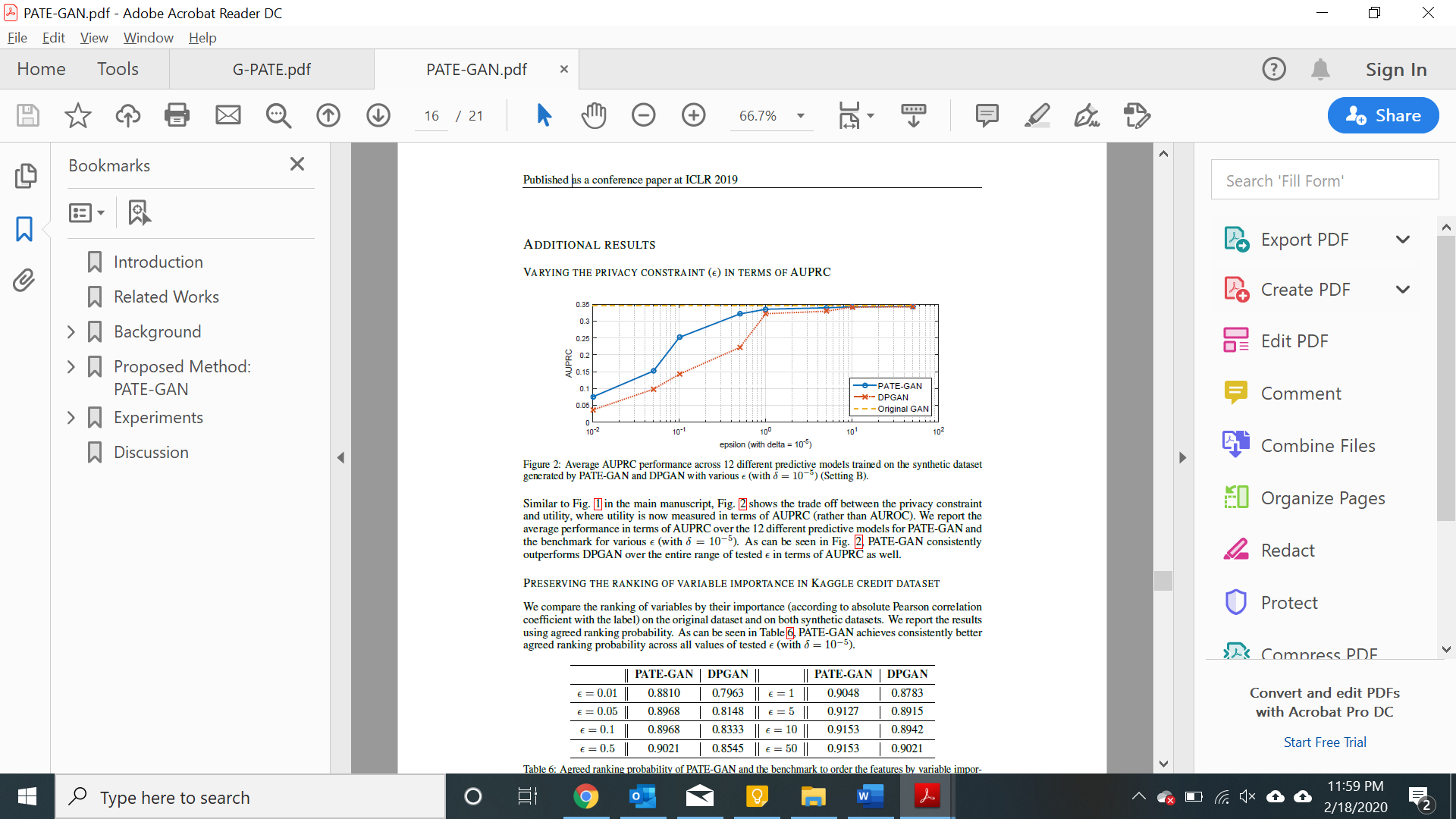
Link: <https://github.com/illidanlab>



**PATE-GAN**

PATE**:** Ensemble of teacher and students can only access teacher output and not teacher data or parameters because of majority voting. Since it is semi-supervised, it assumes that students have access to public dataset. In PATE-GAN, we can use generated data instead of public dataset. Generative Adversarial Networks (GAN) [19] provide a powerful method for using real data to generate synthetic data but it does not provide any rigorous privacy guarantees. Our method modifies the GAN machinery in a way that does guarantee privacy; the synthetic data is (differentially) private [12] with respect to the original data DP-GAN: The key idea is that noise is added to the gradient of the discriminator during training to create differential privacy guarantees. Our method is similar in spirit; during training of the discriminator differentially private training data is used, which results in noisy gradients, however, we use the mechanism introduced in A noticeable difference is that the adversarial training is no longer symmetrical: the teachers are now being trained to improve their loss with respect to G but G is being trained to improve its loss with respect to the student S which in turn is being trained to improve its loss with respect to the teachers.





**G-PATE**

Theoretically, the generator in GAN has the potential of generating an universal distribution, which is a superset of the real distribution, so it is not necessary for the student discriminator to be trained on real records. However, such a theoretical bound is loose. In practice, if a generator does generate enough samples from the universal distribution, there would be a convergence issue. On the other hand, when the generator does converge, it no longer covers the universal distribution, so the student generator may fail to learn the real distribution without seeing real records.

It is not necessary to ensure differential privacy for the discriminator in order to train a differentially private generator. As long as we ensure differential privacy on the information flow from the discriminator to the generator, it is sufficient to guarantee the privacy property for the generator. Therefore, instead of focusing on ensuring differential privacy for the whole GAN framework, we design a novel framework to guarantee that all information flowed from the discriminator to the generator satisfies differential privacy.

Compared to PATE-GAN, our approach has two advantages. First, we improve the use of privacy budget by applying it to the part of the model that actually needs to be released for data generation. Second, our discriminator can be trained on real data because itself does not need to satisfy differential privacy. The teacher discriminators do not need to be published, so they can be trained with non-private algorithms. In addition, we design a gradient aggregator to collect information from teacher discriminators and combine them in a differentially private fashion.

Unlike PATE-GAN, G-PATE does not require any student discriminator. The teacher discriminators are directly connected to the student generator. The gradient aggregator sanitizes the information flow from the teacher discriminators to the student generator to ensure differential privacy The privacy property is achieved by sanitizing all information propagated from the discriminators to the generator.

**T-GAN**

Recently, several GAN models emerged to handle tabular data, especially to generate medical records. RGAN and RCGAN [13] can generate real-valued time-series data. medGAN [9], corrGAN [33] and several improved models [2, 6, 47, 3, 44] can generate discrete medical records but do not tackle the complexity in generating multimodal continuous variables. ehrGAN [8] generates augmented medical records but

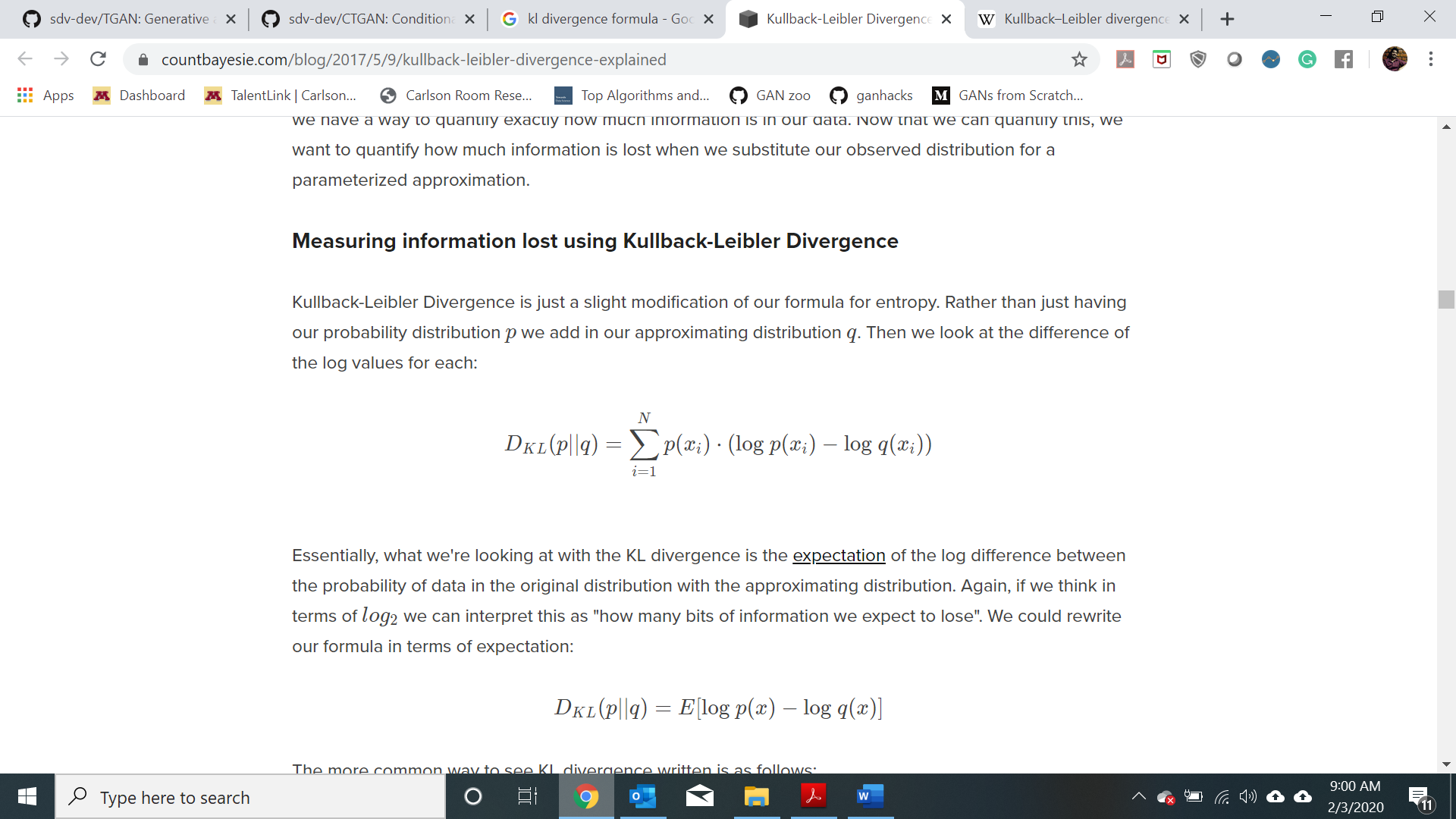
doesn’t explicitly generate synthetic data. Finally, perhaps the work that is closest to our work is tableGAN [31] - that is, it tries to solve the problem of generating synthetic data for a tabular dataset1. However, there are a few fundamental differences. It uses convolutional neural networks while we use recurrent networks. Also, tableGAN explicitly optimizes the prediction accuracy on synthetic data by minimizing cross entropy loss while our model cares more about marginal distribution. We explicitly learn the marginal distribution of each column by minimizing KL divergence.

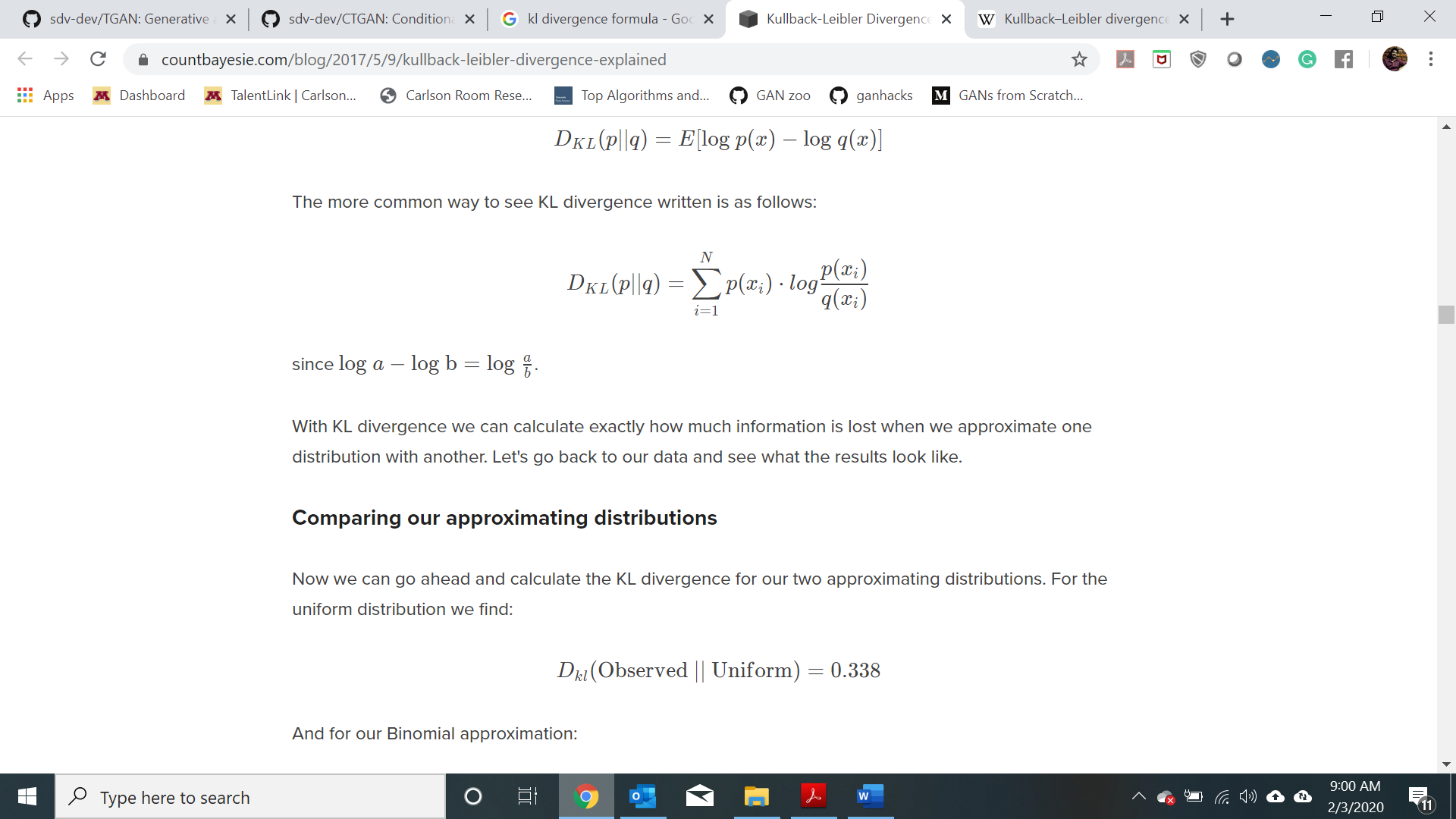
**KL Divergence**

KL divergence is relative difference in probability distribution of Ground truth and prediction.

KL divergence = cross entropy – entropy (ground truth)

So, minimizing KL divergence is equal to minimizing cross entropy because entropy (ground truth) is constant.





**Metrics for similarity:**

1. Likelihood fitness
2. KL divergence – Part of generator model
3. Normalized mutual information for pairwise correlation.
4. Mean, standard deviation, variance

**Metric for re-identification:**

1. Linkage attack